DNA methylation profiling of non-small cell lung cancer reveals a COPD-driven immune-related signature

Els Wauters[1,2,3], Wim Janssens[3], Johan Vansteenkiste[3], Herbert Decaluwé[4], Nele Heulens[5], Bernard Thienpont[1,2], Hui Zhao[1,2], Dominiek Smeets[1,2], Xavier Sagaert[6], Johan Coolen[7], Marc Decramer[3], Adrian Liston[8], Paul De Leyn[4], Matthieu Moisse[1,2,9] and Diether Lambrechts[1,2,9]

1. Vesalius Research Center (VRC), VIB, Leuven, Belgium.
2. Laboratory for Translational Genetics, Department of Oncology, KU Leuven, Leuven, Belgium.
3. Respiratory Division, University Hospital Gasthuisberg, KU Leuven, Leuven, Belgium.
4. Department of Thoracic surgery, University Hospital Gasthuisberg, KU Leuven, Leuven, Belgium.
5. Laboratory of Pneumology, KU Leuven, Leuven, Belgium.
6. Centre for Translational Cell & Tissue Research, KU Leuven, Leuven, Belgium.
7. Department of Radiology, University Hospital Gasthuisberg, KU Leuven, Belgium
8. Autoimmune Genetics Laboratory, VIB, KU Leuven, Leuven, Belgium.
9. Contributed equally to this study

Corresponding author: Diether Lambrechts, MSc, PhD.
Vesalius Research Center (VRC)
VIB and KU Leuven
Campus Gasthuisberg, Herestraat 49, box 912
B-3000, Leuven, Belgium
Tel: +32-16-37.32.09; Fax: +32-16-37.25.85
E-mail: diether.lambrechts@vib-kuleuven.be

Running Title: Lung tumors in COPD versus non-COPD patients differ in immune response

Key words: chronic obstructive pulmonary disease or COPD, lung cancer, methylation, gene expression, immune cell infiltration

Word count: 3,128

KEY MESSAGES

What is the key question?

Although COPD independently increases the risk for NSCLC up to fivefold, it has not been established how COPD influences NSCLC development and tumor progression.

What is the bottom line?

Lung tumors in COPD patients are, in contrast to those from non-COPD patients, characterized by an immune-related signature as revealed by unbiased DNA methylation profiling, and confirmed by gene expression profiling and immunohistochemistry.

Why read on?

The identified immune-related signature in COPD-associated NSCLC may contribute to the optimization of novel NSCLC treatment strategies by assisting in treatment stratification and prediction of treatment response.

ABSTRACT

Introduction Non-small cell lung cancer (NSCLC) is a heterogeneous disorder consisting of distinct molecular subtypes each characterized by specific genetic and epigenetic profiles. Here, we aimed to identify novel NSCLC subtypes based on genome-wide methylation data, assess their relationship with smoking behavior, age, chronic obstructive pulmonary disease (COPD), emphysema and tumor histopathology, and identify the molecular pathways underlying each subtype.

Methods Methylation profiling was performed on 49 pairs of tumor and adjacent lung tissue using Illumina 450K arrays. Transcriptome sequencing was performed using Illumina HiSeq2000 and validated using expression data from The Cancer Genome Atlas (TCGA). Tumor immune cell infiltration was investigated by immunohistochemistry.

Results Unsupervised hierarchical clustering of tumor methylation data revealed 2 subgroups characterized by a significant association between cluster membership and presence of COPD (P=0.024). Ontology analysis of genes containing differentially methylated CpGs (FDR-adjusted P<0.05) revealed that immune genes were strongly enriched in COPD tumors, but not in non-COPD tumors. This COPD-specific immune signature was attributable to methylation changes in immune genes expressed either by tumor cells or tumor-infiltrating immune cells. No such differences were observed in adjacent tissue. Transcriptome profiling similarly revealed that genes involved in the immune response were differentially expressed in COPD tumors (FDR-adjusted P<0.05), an observation that was independently replicated using TCGA data. Immunohistochemistry validated these findings, revealing fewer CD4-positive T-lymphocytes in tumors derived from COPD patients.

Conclusions Lung tumors of COPD patients differ from those of non-COPD patients, with differentially methylated and expressed genes being mainly involved in the immune response.

INTRODUCTION

Lung cancer is the leading cause of cancer-related death worldwide with an alarmingly low 5-year survival rate of <15%. Therapies targeting specific genetic lesions in the tumor are increasingly being used to treat lung cancer, highlighting the need to accurately classify lung tumors based on their molecular-genetic composition.[1] Similar to somatic mutations, epigenetic modifications may provide a substrate for clonal selection during malignant transformation. As such, genome-wide methylation profiling studies can identify signatures that contribute to the molecular sub-classification of various cancers.[2] For instance, methylation profiling classifies breast tumors into its four established expression subtypes and already identified an additional subtype associated with T-lymphocyte infiltration and favorable outcome.[3]

Over 50% of lung cancer patients also suffer from chronic obstructive pulmonary disease (COPD) due to shared risk factors, such as age and cigarette smoking.[4] However, COPD also independently increases the risk for lung cancer up to fivefold.[5] Although genetic association studies revealed susceptibility variants overlapping between both diseases,[6,7] the precise mechanisms underlying this link are incompletely understood.[8] Here, we hypothesize that the inflammatory microenvironment in COPD lungs, which results from a maladaptive response of the immune system to inhaled noxious particles, may influence lung cancer development through its impact on the epigenome. Indeed, amongst various biological processes, changes in immune response-related genes were recently observed in peripheral blood leucocytes and bronchial small airway epithelial cells of cancer-free smokers with COPD.[9] Likewise, age-related CpG island hypermethylation in non-malignant human tissues and gene-specific hypomethylation in blood of active smokers have been

described.[10] Surprisingly, although lung cancer-specific methylation profiles have previously been generated, the potential contributions of COPD, age and smoking history to these signatures have not been assessed.

METHODS

Additional details can be found in the online supplement.

Study participants

We prospectively recruited untreated NSCLC patients scheduled for curative-intent surgery and a smoking history of ≥15 pack-years at the Leuven University Hospital (Belgium) between March 2010 and August 2011. Prior to surgery, all 49 eligible participants performed spirometry according to the American Thoracic Society and European Respiratory Society guidelines.[11] COPD was defined as post-bronchodilator $FEV_1/FVC<0.70$ and staged according GOLD 2007 guidelines.[12] Emphysema was semi-quantitatively assessed by a radiologist, as described previously.[6,7] Other respiratory diseases affecting pulmonary function were excluded. Previous treatments for COPD were retrieved from electronic medical files. The study protocol was approved by the local ethical committee. All participants provided written informed consent.

DNA isolation and methylation profiling

DNA was extracted from fresh-frozen tumor tissue, in which ≥70% of the epithelial cells were malignant based on histological examination, and adjacent lung tissue from the same resection specimen at maximal distance (>5cm) from the tumor. Bisulfite-converted DNA was hybridized on 450K BeadChips, using the Infinium HD Methylation protocol. After controlling for batch effect and normalizing raw intensity data, the methylation level of each interrogated CpG, scored as a β-value between 0

and 1, was calculated according to the fluorescent intensity ratio.

RNA extraction and transcriptome sequencing

Total RNA was extracted from all tumor samples undergoing methylation profiling. RNA sequencing was performed using Illumina HiSeq2000. In addition, RNA sequencing data of tumor and adjacent lung tissue samples of 178 NSCLC patients were downloaded from the TCGA portal (https://tcga-data.nci.nih.gov/tcga/).

Histological analyses

An experienced pathologist blinded to the clinical data semi-quantitatively scored hematoxylin and eosin (H&E)-stained tumor sections for immune cell infiltrates, as no-to-mild or moderate-to-severe infiltration present. Consecutive tumor sections were stained with antibodies against human T-cells (CD3, CD4, CD8), B-cells (CD20), macrophages (CD68), neutrophils (MPO, CD66b) and natural killer cells (NCAM). Stained sections were screened at low power (100x) to identify 5 areas with the highest density of positive cells (hot spots). Next, at each hot spot, an image was taken at 400x magnification (high power field, HPF) and the absolute number of positive cells per hotspot surface was counted manually to obtain the average density of positive cells per 5 HPFs (positive cells/mm^2).

Statistical analyses

Methylation analyses were performed on normalized β-values. Top-down unsupervised hierarchical clustering was performed using a recursively partitioned mixture model (RPMM).[13] Multivariate logistic regression models with forward stepwise selection assessed age, histological subtype, pack-years smoked, years-quit smoking, COPD status, emphysema, COPD treatment and batch as independent predictors. Differential methylation was assessed between tumor and adjacent lung tissue in the entire study population, in COPD and non-COPD subgroups. We used a

paired linear model for differential methylation assessment (Limma). A CpG site was differentially methylated when exhibiting a false discovery rate (FDR)-adjusted *P*-value <0.05 and a difference in β-value between tumor and adjacent lung tissue (Δβ-value) >0.075.[14] Differential expression analysis was performed using DESeq following normalization with EDASeq. A two-fold or greater change in expression and FDR<0.05 were selected to designate significantly up- or downregulated genes. Gene ontology (GO) enrichment analysis was performed using GOrilla. Only GO-terms with a FDR<0.05 were considered as significantly enriched. The association of COPD with the number of tumor-infiltrating immune cells was assessed by linear regression adjusted for covariates. Statistical analyses were performed using SPSS software (v19.0) and R (v2.14.1).

RESULTS

Novel NSCLC subtypes based on DNA methylation profiling

We included 49 NSCLC patients that underwent curative-intent surgery. Tissue pairs from 3 patients were excluded due to low signal intensities for some methylation probes, as described in the online supplement. Clinical characteristics of the remaining 46 patients are shown in table 1.

Pairwise comparison of methylation values between tumor and adjacent lung tissue (Δβ-values) revealed tumor-associated methylation changes similar to those previously described (see online supplementary results and figure S1).[15] For instance, tumor-associated hypermethylation occurred 3 times more frequently in gene promoters targeted by Polycomb group proteins than in other promoters (i.e., 26.9% versus 8.8%, *P*<0.001).[16]

To subsequently identify novel NSCLC subtypes, unsupervised RPMM-based clustering of all tumor β-values that passed quality control was performed, revealing two major clusters (figure 1; online supplementary results and table S1). Forward-logistic regression established an independent association between tumor cluster membership and COPD status (P=0.024), but not with other covariates in the regression model (age, histological subtype, pack-years smoked, years-quit smoking, emphysema, COPD treatment and batch). More specifically, we found that one cluster was significantly enriched in COPD patients (15/19 or 78.9% versus 12/27 or 44.4%; table 1). Similarly, clustering of β-values in adjacent lung tissue also revealed two clusters, one of which was enriched in COPD patients (78.6% versus 50.0%). COPD status did, however, not significantly influence cluster membership, as assessed by forward-logistic regression (online supplementary table S2). There was no association between other clinical variables and cluster membership. Additionally, no overlap in cluster membership between the tumor- and adjacent tissue-based clusters was noted.

Differential methylation according to COPD status

Since COPD status differed significantly between cluster subgroups, we stratified our study population by COPD status. Clinical characteristics of COPD and non-COPD subgroups are presented in online supplementary table S3. First, to identify COPD-associated methylation events, we compared for each CpG the β-values between tumor versus adjacent lung tissue in the COPD and non-COPD subgroup separately. Subsequently, we selected for both subgroups those CpGs that displayed a significant difference in Δβ-value ($\Delta\beta>0.075$ and FDR<0.05 using a paired linear model; online supplementary table S1) and performed a GO enrichment analysis. In the COPD subgroup, genes with CpGs that were hypermethylated in tumor versus adjacent tissue were enriched for GO terms involved in immune response activation, whereas no

enrichment was observed in the non-COPD subgroup (figure 2A,C). On the other hand, genes with hypomethylated CpGs were not enriched for GO terms in the COPD subgroup, whereas a diverse set of GO processes was identified in the non-COPD subgroup (figure 2B,D).

To assess whether the enrichment for immune response genes in COPD patients originated from tumor or adjacent lung tissue, we directly compared tumor β-values between patients with and without COPD. GO analysis of genes relatively hypermethylated in COPD tumors ($P<0.05$; online supplementary table S1) confirmed that these genes were significantly enriched for immune genes. When comparing adjacent lung tissue between both subgroups, no enrichment for GO terms was found, which is in accordance with the less-pronounced differences in methylation-based clusters in adjacent lung tissue. Hence, hypermethylation of immune response genes in COPD patients specifically affects the tumor tissue.

In a next step, we assessed whether the immune signature in COPD tumor tissue was attributable to a difference in the amount of immune cells infiltrating the tumor or to differential methylation of immune-related genes expressed by tumor cells. We found evidence for both mechanisms. In particular, using the Immunological Genome Project database,[17] we found that genes specifically expressed by immune cells, such as *CD4*, which is expressed by mature T-helper cells, exhibited promoter hypermethylation in COPD tumors. Since promoters of expressed genes are usually not methylated, their high methylation status in COPD tumors suggests that there is reduced infiltration of immune cells in these tumors. For other immune genes, we found published evidence that they are predominantly expressed in tumor cells and that changes in their expression levels influence the anti-tumor immune response. These genes included *CCL5*, encoding a chemo-attractant;

TNFRSF21, encoding a member of the TNF receptor superfamily; and *SUSD2*, influencing tumor infiltration by CD4-positive lymphocytes. Hypermethylation of the latter genes in COPD patients is therefore likely to occur in tumor cells.

Methylation-driven gene expression changes

Next, to investigate the impact of tumor-associated DNA hypermethylation on gene expression, we sequenced mRNA of 39 tumor-normal pairs (other pairs failed RNA quality control assessment; see online supplementary results). On average, 36.0 million sequence reads were generated per sample and 69.4% of coding genes were expressed.

Given the observed differences in methylation between tumors in COPD and non-COPD patients, we stratified expression data for COPD status (21 and 18 patients with and without COPD, respectively; online supplementary table S4). Moreover, since immune response genes were hypermethylated in COPD tumors and DNA hypermethylation of gene promoters results in silencing of the associated gene, we focused on genes downregulated in tumor versus adjacent tissue (FDR<0.05; online supplementary table S1). GO-analysis of genes characterized by reduced expression in the tumor revealed a strong enrichment of immune genes in the COPD subgroup only, confirming that the observed hypermethylation events in COPD tumors were relevant at the expression level (figure 3A,B). Notably, also after correcting for COPD treatment, GO analysis revealed the same enrichment for immune genes. In adjacent tissue, gene expression changes did not show enrichment for immune-related GO terms in COPD versus non-COPD patients, although a significant enrichment in immune-related GO terms among genes significantly downregulated (P<0.05) was noted in adjacent tissue from treated versus non-treated COPD patients (online supplementary figure S2), as reported previously.[18] However, since a less stringent *P*-

value threshold was used to select downregulated genes in this analysis, this effect is clearly less pronounced than the immune-related signature in COPD tumors.

Next, we integrated DNA methylation and RNA sequencing data obtained from COPD patients (online supplementary figure S3). Since GO terms were not enriched in hypomethylated genes when comparing COPD tumor versus adjacent lung tissue, we focused on hypermethylated genes. Of the 1,985 genes with promoter hypermethylation in tumor versus adjacent tissue (FDR<0.05), we identified 311 genes whose expression was more than two-fold downregulated in tumor versus adjacent tissue of COPD patients (FDR<0.05; online supplementary table S1). These contained 76 immune-related genes (figure 4A,B).

Validation of an immune-specific signature in lung cancer patients with COPD

As an independent validation of the COPD-specific expression signature, we used TCGA expression data of 178 NSCLC patients for whom spirometry data were available.[19] After stratification of this cohort for COPD status (46 patients with and 132 patients without COPD; online supplementary table S5), GO analysis of genes downregulated in tumor versus adjacent tissue (FDR<0.05; online supplementary table S1) confirmed the enrichment of immune genes in the COPD subgroup, thus independently validating the presence of an immune-specific signal in COPD tumors (figure 3C,D).

Finally, we also assessed which immune genes were downregulated in tumor versus adjacent tissue from COPD patients both in our discovery and validation cohort (FDR<0.05 in both cohorts). Remarkably, of the 535 genes that were significantly downregulated in COPD tumors in both cohorts, 161 genes were associated with an immune-related GO term (online supplementary table S1). Interestingly, of the 311 genes that were both hypermethylated and downregulated in

our discovery cohort, 125 genes were also downregulated in tumor versus adjacent tissue in the validation cohort (FDR<0.05). Genes were involved in the innate defense response (*AGER*), lung dendritic cell trafficking (*CCRL2*), lymphocyte migration (*S1PR1*), etc. (figure 5).

Histopathology of immune cell infiltrates in COPD versus non-COPD tumors

We also quantified infiltrating immune cells using histopathology. First, an experienced pathologist estimated the degree of immune cell infiltration on H&E-stained sections from all 49 tumors selected for methylation profiling. This revealed significantly less immune cell infiltration in COPD patients (*P*=0.036; figure 6A-C; online supplementary figure S4). Multivariable-adjusted logistic regression confirmed that this effect was independent from covariates (*P*=0.022).

Secondly, we stained consecutive tumor sections for established T-cell (CD3, CD4, CD8), B-cell (CD20), neutrophil (MPO, CD66b), macrophage (CD68) and natural killer cell (NCAM) markers. Multivariable-adjusted linear regression revealed a significant association between COPD status and the average density of CD3-positive T-cells (*P*=0.001). In particular, tumors from COPD patients contained fewer CD3-positive T-cells (figure 6D-F). There was a similar correlation between COPD status and the density of CD4-positive cells (*P*=0.012; figure 6G-I). CD4 is commonly used as a T-helper (T_H) cell marker. However, it can also be expressed by other immune cells, such as macrophages. The fact that we observed reduced infiltration for both CD3- and CD4-positive cells suggests that infiltration of CD4-positive T_H-cells in COPD tumors was reduced. This was confirmed microscopically based on typical T_H-cell morphology. Furthermore, there were no significant differences observed for the CD68 macrophage-specific marker (online supplementary figure S5).

Finally, to elucidate whether specific T_H-cell subtypes were affected, we re-analyzed gene expression data from the discovery cohort. In particular, by assessing T_H-cell subtypes (T_H1, T_H2, T_H9, T_H17, T_H22, regulatory (T_{Reg}) and follicular (T_{FH}) T-cells) using a published marker set[20] (online supplementary table S6), we observed that global expression of T_H-cell markers in COPD tumors was indeed reduced (P=0.018). Particularly, at the individual T_H-cell subtype level, expression of T_{Reg} markers was significantly reduced in COPD tumors (P=0.046, online supplementary figure S6).

DISCUSSION

We studied 49 pairs of surgically-resected tumors and adjacent tissues using genome-wide methylation profiling. Unsupervised clustering of tumor data identified two subgroups that represented different populations in terms of COPD status. Although other clinical covariates are also likely to affect the methylation profile of the tumor, our data suggest that the effect of COPD is stronger than that of age, smoking behavior, emphysema and histopathological subtype. This is surprising, because age and smoking are major risk factors for NSCLC, and because histopathology reflects the difference in cell type affected by the carcinogenic process, whereby cell type identity is a well-known determinant of DNA methylation. In the adjacent tissue, no such difference was observed in COPD versus non-COPD patients, thereby confirming previous DNA methylation and gene expression studies.[21-26] Only when stratifying for COPD treatment, a weak enrichment in immune-related GO terms was observed for genes downregulated in the adjacent tissue of the treated subgroup of COPD patients, as reported by others.[18] COPD thus seems to more strongly affect DNA methylation of tumor than of normal tissue.

Stratification for COPD status revealed that epigenetic events in COPD-associated NSCLC were enriched for immune-related genes. This was attributable to a relative hypermethylation of immune genes either predominantly expressed by tumor-infiltrating immune cells or by tumor cells. Promoters of genes expressed by immune cells are mostly unmethylated, but may become hypermethylated (and thus not expressed) in tumor cells. A higher methylation status of such genes thus suggests that there is reduced infiltration of immune cells in the tumor. This was indeed confirmed by immunohistopathology. Combined with our observation that also immune genes predominantly expressed by tumor cells were hypermethylated, these findings suggest that COPD tumors are characterized by a less pronounced immune response.

Since COPD lungs are characterized by increased immune cell infiltration and excessive chronic inflammation,[27] these findings may look surprising. However, often the immune system of COPD patients is largely dysfunctional. Their macrophages, for instance, exhibit deficient phagocytic functions towards apoptotic epithelial cells and bacteria, thereby sustaining chronic inflammation and predisposing to infections that are characteristic of severe COPD.[27-30] The immune system in COPD patients may thus not be capable of mounting an effective immune response against the tumor. As a result, in COPD tumors, the infiltration of T_H-cells will be less pronounced, which will further weaken anti-tumor immunity, as T_H-cells are essential for inducing a strong immune response. Accordingly, the more pronounced downregulation of markers for T_{Reg}-cells, which in principle have a tumor-associated immunosuppressive effect, supports the hypothesis that COPD tumors are less dependent on immunosuppression.

Overall, we thus identified a NSCLC subtype characterized by reduced immune cell infiltration and associated with COPD status. Interestingly, cancer subtypes characterized by differential expression of immune response genes have previously been established as predictors of treatment response and survival in other cancers.[31-33] Future studies assessing novel NSCLC treatment strategies should thus take COPD status or reduced immune cell infiltration as a potential predictive biomarker into account. Especially when testing cancer immunotherapies, COPD status might have to be considered. Very recently, the FDA approved the anti-PD-1 therapy, nivolumab, as the first immunotherapy for metastatic NSCLC.[34] It would obviously be very interesting to assess whether NSCLC patients respond differently to this novel therapy depending on their COPD status.

Although methylation profiling studies have previously been conducted for NSCLC tumors, our study is unique in several respects. First of all, our patient population was extensively phenotyped and for the first time included COPD status as a covariate (online supplementary table S7). Secondly, adjacent non-malignant lung tissue was collected for each tumor sample, allowing us to correct for inter-individual epigenetic variation due to aging and smoking. Indeed, immune-related GO terms were more enriched when analyzing Δβ-values than tumor β-values. A potential limitation is that there were no methylation data publically available to replicate our clustering results. Additionally, as it is virtually impossible to collect lung tissue from cancer-free non-COPD patients (because these patients do not undergo surgery), we were forced to compare COPD versus non-COPD adjacent lung tissue from NSCLC patients undergoing surgery. Effects of the tumor on the adjacent resection specimen can thus not formally be excluded. Finally, systemic corticosteroid use has been shown to influence DNA methylation in COPD patients.[35] However, although other

COPD treatments also affected gene expression, we were not able to explore this specifically for systemic corticosteroids given the limited number of patients treated (n=4).

Overall, our study describes the presence of a specific tumor micro-environment in COPD-associated NSCLC, which is characterized by reduced immune cell infiltration, thereby providing novel insights into how lung tumors of COPD patients differ from those of non-COPD patients.

Acknowledgments The authors thank Thomas Van Brussel, Gilian Peuteman, Stephan Vinckier, Kristien De Bent, Maarten Spruyt, Erica Balligand, An Lehouck and Beatrijs Anrijs for their excellent assistance during data collection and for technical support. We also thank the lung function technicians from the UZ Leuven for their dedication and efforts.

Contributors Conception and design: WJ, JV and DL; Experimental work and data analysis: EW, MM, HD, NH, HZ, JC, DS, XS and PDL; Data interpretation and manuscript writing: EW, BT, WJ, JV, MD, AL, MM and DL. All authors read and approved the final version of the manuscript.

Funding EW, MM, HZ and BT are supported by the Fund for Scientific Research Flanders (FWO-F). MM is supported by SymBioSys. WJ is clinical researcher of FWO-F. This work was supported by a European Research Council (ERC) consolidator grant RCN:191995 (DL), by a FWO-F research project G065615N and by Vlaamse Liga tegen Kanker (VLK).

Competing interests None of the authors has a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Ethics approval Ethical Comity University Hospital Gasthuisberg.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Booton, R., Blackhall, F., Kerr, K. Individualised treatment in non-small cell lung cancer: precise tissue diagnosis for all? *Thorax* 2011;66:273-5.

2 Heyn, H., Esteller, M. DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet* 2012;13:679-92.

3 Dedeurwaerder, S., Desmedt, C., Calonne, E. *et al.* DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol Med* 2011;3:726-41.

4 Young, R. P., Hopkins, R. J., Christmas, T. *et al.* COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. *Eur Respir J* 2009;34:380-6.

5 Mannino, D. M., Aguayo, S. M., Petty, T. L. *et al.* Low lung function and incident lung cancer in the United States: data From the First National Health and Nutrition Examination Survey follow-up. *Arch Intern Med* 2003;163:1475-80.

6 Lambrechts, D., Buysschaert, I., Zanen, P. *et al.* The 15q24/25 susceptibility variant for lung cancer and chronic obstructive pulmonary disease is associated with emphysema. *Am J Respir Crit Care Med* 2010;181:486-93.

7 Wauters, E., Smeets, D., Coolen, J. *et al.* The TERT-CLPTM1L locus for lung cancer predisposes to bronchial obstruction and emphysema. *Eur Respir J* 2011;38:924-31.

8 Houghton, A. M. Mechanistic links between COPD and lung cancer. *Nat Rev Cancer* 2013;13:233-45.

9 Vucic, E. A., Chari, R., Thu, K. L. *et al.* DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways. *Am J Respir Cell Mol Biol* 2014;50:912-22.

10 Christensen, B. C., Houseman, E. A., Marsit, C. J. *et al.* Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 2009;5:e1000602.

11 Celli, B. R., MacNee, W. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004;23:932-46.

12 Rabe, K. F., Hurd, S., Anzueto, A. *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007;176:532-55.

13 Houseman, E. A., Christensen, B. C., Yeh, R. F. *et al.* Model-based clustering of DNA methylation array data: a recursive-partitioning algorithm for high-dimensional data arising as a mixture of beta distributions. *BMC bioinformatics* 2008;9:365.

14 Benjamini, Y., Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Statist Soc* 1995;57:289-300.

15 Easwaran, H., Johnstone, S. E., Van Neste, L. *et al.* A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res* 2012;22:837-49.

16 Ku, M., Koche, R. P., Rheinbay, E. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 2008;4:e1000242.

17 Benoist, C., Lanier, L., Merad, M. *et al.* Consortium biology in immunology: the perspective from the Immunological Genome Project. *Nat Rev Immunol* 2012;12:734-40.

18 van den Berge, M., Steiling, K., Timens, W. *et al.* Airway gene expression in COPD is dynamic with inhaled corticosteroid treatment and reflects biological pathways associated with disease activity. *Thorax* 2014;69:14-23.

19 Weinstein, J. N., Collisson, E. A., Mills, G. B. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013;45:1113-20.

20 Tangye, S. G., Ma, C. S., Brink, R. *et al.* The good, the bad and the ugly - TFH cells in human health and disease. *Nat Rev Immunol* 2013;13:412-26.

21 Bhattacharya, S., Srisuma, S., Demeo, D. L. *et al.* Molecular biomarkers for quantitative and discrete COPD phenotypes. *Am J Respir Cell Mol Biol* 2009;40:359-67.

22 Lamontagne, M., Timens, W., Hao, K. *et al.* Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction. *Thorax* 2014;69:997-1004.

23 Yoo, S., Takikawa, S., Geraghty, P. *et al.* Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. *PLoS Genet* 2015;11:e1004898.

24 Sato, T., Arai, E., Kohno, T. *et al.* Epigenetic clustering of lung adenocarcinomas based on DNA methylation profiles in adjacent lung tissue: Its correlation with smoking history and chronic obstructive pulmonary disease. *Int J Cancer* 2014;135:319-34.

25 Kim, W. J., Lim, J. H., Lee, J. S. *et al.* Comprehensive Analysis of Transcriptome Sequencing Data in the Lung Tissues of COPD Subjects. *Int J Genomics* 2015;2015:206937.

26 Wang, I. M., Stepaniants, S., Boie, Y. *et al.* Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. *Am J Respir Crit Care Med* 2008;177:402-11.

27 Brusselle, G. G., Joos, G. F., Bracke, K. R. New insights into the immunology of chronic obstructive pulmonary disease. *Lancet* 2011;378:1015-26.

28 Berenson, C. S., Kruzel, R. L., Eberhardt, E. *et al.* Impaired innate immune alveolar macrophage response and the predilection for COPD exacerbations. *Thorax* 2014;69:811-8.

29 Dancer, R., Sansom, D. M. Regulatory T cells and COPD. *Thorax* 2013;68:1176-8.

30 Hou, J., Sun, Y., Hao, Y. *et al.* Imbalance between subpopulations of regulatory T cells in COPD. *Thorax* 2013;68:1131-9.

31 Budinska, E., Popovici, V., Tejpar, S. *et al.* Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J Pathol* 2013;231:63-76.

32 Tan, T. Z., Miow, Q. H., Huang, R. Y. *et al.* Functional genomics identifies five distinct molecular subtypes with clinical relevance and pathways for growth control in epithelial ovarian cancer. *EMBO Mol Med* 2013;5:983-98.

33 Ignatiadis, M., Singhal, S. K., Desmedt, C. *et al.* Gene modules and response to neoadjuvant chemotherapy in breast cancer subtypes: a pooled analysis. *J Clin Oncol* 2012;30:1996-2004.

34 Rizvi, N. A., Mazieres, J., Planchard, D. *et al.* Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced,

refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial. *Lancet Oncol* 2015;16:257-65.

35 Wan, E. S., Qiu, W., Baccarelli, A. *et al.* Systemic steroid exposure is associated with differential methylation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;186:1248-55.

FIGURE LEGENDS

Figure 1. Unsupervised clustering of tumor β-values and heatmap representation of DNA methylation data.

Left panel: RPMM-based classification of 46 NSCLC tumor samples (columns) and heatmap representation of the 13,965 most variable β-values (rows) between the two clusters (FDR<0.01 and Δβ-value between both clusters >0.15). β-values are represented by a color scale ranging from dark blue (low DNA methylation) to bright yellow (high DNA methylation). Clustering revealed two tumor clusters (n=27 and n=19 for cluster 1 and 2, respectively), which differed according to COPD status, as indicated above the heatmap. Samples from COPD patients are indicated in red above the heatmap, adenocarcinoma samples are indicated in blue and patients receiving COPD treatment are indicated in orange. Continuous variables (age, pack-years smoked and years-quit smoking) and the degree of emphysema are represented by color scales (respectively, black, purple, green and turquoise) ranging from light (low number) to dark (high number). *Right panel:* For comparison, a heatmap representation of the 13,965 β-values identified by clustering of tumor β-values (left panel) in the 46 adjacent non-malignant lung tissues is shown. All adjacent lung tissue samples have comparable methylation values for the selected tumor-associated CpGs.

Figure 2. GO terms enriched in genes differentially methylated in NSCLC patients with and without COPD.

GO enrichment analysis was performed on genes containing significantly differentially methylated β-values in tumor versus adjacent tissue in the COPD subgroup (*left panel*) or in the non-COPD subgroup (*right panel*). *Left panel:* For the

COPD subgroup: (A) the 5 most significantly enriched GO terms among the 546 hypermethylated genes are presented; (B) no GO enrichment was identified among the 2,154 hypomethylated genes. *Right panel:* For the non-COPD subgroup: (C) no GO enrichment was identified among the 1,835 hypermethylated genes; (D) the 5 most significantly enriched GO terms among the 1,169 hypomethylated genes are presented. Horizontal bars represent the fold enrichment (X-axis) for the significantly enriched GO terms, and bars for GO terms associated with immunological processes are red. The FDR-values for enrichment, the number of differentially methylated genes associated with a specific GO term (n), and the ratio of n and the total number of genes associated with a specific GO term (%) are provided at the right of the associated bars.

Figure 3. GO terms enriched in genes differentially expressed in NSCLC patients with and without COPD.

GO enrichment analysis was performed on genes significantly downregulated in tumor versus adjacent lung tissue from NSCLC patients with COPD (n=445 genes, A) and without COPD (n=295 genes, B) in our RNA-seq discovery cohort (*left panel*); and on genes significantly downregulated in tumor versus adjacent lung tissue from NSCLC patients with COPD (n=34 genes, C) and without COPD (n=698 genes, D) in the RNA-seq TCGA validation cohort (*right panel*). The 5 most significantly enriched GO terms for each subgroup are presented. Horizontal bars represent the fold enrichment (X-axis) for the significantly enriched GO terms, and bars for GO terms associated with immunological processes are red. The FDR-values for enrichment, the number of differentially expressed genes associated with a specific GO term (n), and

the ratio of n and the total number of genes associated with a specific GO term (%) are provided at the right of the associated bars.

Figure 4. Integration of DNA methylation and gene expression changes.

Box plots showing changes in promoter DNA methylation (A) and gene expression (B) between tumor tissue (grey boxes) and adjacent lung tissue (white boxes) from NSCLC patients with COPD for 12 immune genes found to be both promoter hypermethylated and downregulated, including *CD157* involved in migration of neutrophils to the inflammation site and *CCRL2* involved in lung dendritic cell trafficking and priming of a T-helper cell response. The Y-axis represents the mean promoter β-value (A) and the gene expression as the $\log_2$-transformed fold change over median gene expression level for each sample (B) for the selected genes. The box represents the interquartile ranges. The line across the box indicates the median. The whiskers extend from the box to the 90th and 10th percentiles, excluding outliers. Outliers (open circles) are defined as cases outside the 90th and 10th percentiles. For all selected genes, both methylation and gene expression differed significantly between tumor and adjacent lung tissue from COPD patients (FDR-adjusted *P*-values <0.05).

Figure 5. Expression of selected immune markers downregulated in the RNA-seq discovery and validation cohort.

Box plots showing gene expression levels of 12 immune genes with a significantly lower expression level in tumor tissue from COPD patients (grey boxes) versus tumor tissue from non-COPD patients (white boxes), both in the discovery and TCGA validation cohort (FDR<0.05 for both). The Y-axis represents the gene expression as

the log_2-transformed fold change over median gene expression level for each sample. The box represents the interquartile ranges.

Figure 6. Fewer tumor-infiltrating immune cells in COPD versus non-COPD patients.

(*Top panel*) Representative microscopic pictures showing different degrees of immune cell infiltration in H&E stained tumor sections. (A) Tumor section with mainly fibrosis (indicated by the white arrow) and a low density of immune cell infiltrates between tumor cell nests (some of which are marked by a white asterisk). This section is representative for tumor sections with no-to-mild immune cell infiltration. (B) Tumor section with pronounced immune cell infiltration (indicated by the white arrow) between the tumor cell nests (some of which are marked by a white asterisk). This section is representative for tumor sections with moderate-to-severe immune cell infiltration. (C) Stacked bar plot showing the proportion of tumors showing no-to-mild versus moderate-to-severe (colored in grey) immune cell infiltration between COPD and non-COPD patients. A significantly higher proportion of COPD tumors showed no-to-mild immune cell infiltration.

(*Middle panel*) Representative pictures of tumor sections from COPD (D) and non-COPD patients (E) stained with an antibody directed against CD3. Positive cells are colored brown. (F) Dot chart showing the density of tumor-infiltrating CD3-positive cells in COPD versus non-COPD patients. The Y-axis represents the average density of positive cells per 5 high power fields (HPFs). The horizontal black bar represents the mean.

(*Lower panel*). Representative pictures of tumor sections from COPD (G) and non-COPD patients (H) stained with an antibody directed against CD4. Positive cells are colored brown. (I) Dot chart showing the density of tumor-infiltrating CD4-positive

cells in COPD versus non-COPD patients. The Y-axis represents the average density of positive cells per 5 high power fields (HPFs). The horizontal black bar represents the mean. Asterisks (*) indicate $P<0.05$. Original magnifications: x50 (A,B) and x400 (D,E,G,H). Scale bar, 200μm (upper panel) and 50μm (middle and lower panel).

TABLE 1. Clinical characteristics of the study population and of DNA methylation-based tumor clusters

	Total	Cluster 1	Cluster 2
Subjects, n (%)	46	27 (58.7)	19 (41.3)
Demographics			
Age, years, mean (SD)	67.0 (8.0)	66.1 (8.4)	68.3 (7.4)
Male sex, n (%)	39 (84.8)	23 (85.2)	16 (84.2)
Smoking behavior			
Pack-years history, mean (SD)	44.3 (23.4)	44.4 (27.0)	44.1 (17.8)
Smoked years, mean (SD)	43.0 (12.7)	40.4 (13.2)	46.8 (11.2)
Current smokers, n (%)	29 (63.0)	16 (59.3)	13 (68.4)
COPD, n (%)	27 (58.7)	12 (44.4)	15 (78.9)
COPD GOLD stage, n (%)			
stage I	11 (23.9)	4 (33.3)	7 (46.7)
stage II	15 (32.6)	8 (66.7)	7 (46.7)
stage III	0 (0.0)	0 (0.0)	0 (0.0)
stage IV	1 (2.2)	0 (0.0)	1 (6.7)
Histologic type, n (%)			
Adenocarcinoma	25 (54.3)	16 (59.3)	9 (47.4)
Squamous cell	21 (45.7)	11 (40.7)	10 (52.6)
Tumor stage, n (%)			
IA	11 (23.9)	4 (14.8)	7 (36.8)
IB	9 (19.6)	5 (18.5)	4 (21.1)
IIA	9 (19.6)	6 (22.2)	3 (15.8)
IIB	6 (13.0)	5 (18.5)	1 (5.3)
IIIA	9 (19.6)	6 (22.2)	3 (15.8)
IV	2 (4.3)	1 (3.7)	1 (5.3)
COPD Treatment, n (%)	16 (34.8)	9 (33.3)	7 (36.8)
LAMA	7 (15.2)	5 (18.5)	2 (10.5)
LABA	1 (2.2)	1 (3.7)	0 (0.0)
ICS	14 (30.4)	8 (29.6)	6 (31.6)
CS	4 (8.7)	3 (11.1)	1 (5.3)
Emphysema, n (%)			
None	10 (21.7)	5 (18.5)	5 (26.3)
Mild	27 (58.7)	17 (63)	10 (52.6)
Moderate	8 (17.4)	5 (18.5)	3 (15.8)
Severe	1 (2.2)	0 (0.0)	1 (5.3)

Categorical variables are expressed as number (n) and percentage (%). Percentages represent column percentages. Continuous variables are expressed as mean and standard deviation (SD). Abbreviations: COPD, Chronic Obstructive Pulmonary Disease; GOLD, Global Initiative for Chronic Obstructive Lung Disease; LAMA, long-acting muscarinic antagonists; LABA, inhaled long-acting β 2-agonists; ICS, inhaled corticosteroid; CS, oral corticosteroids. Clusters were determined by RPMM-based clustering of normalized tumor β-values.

Figure 1.

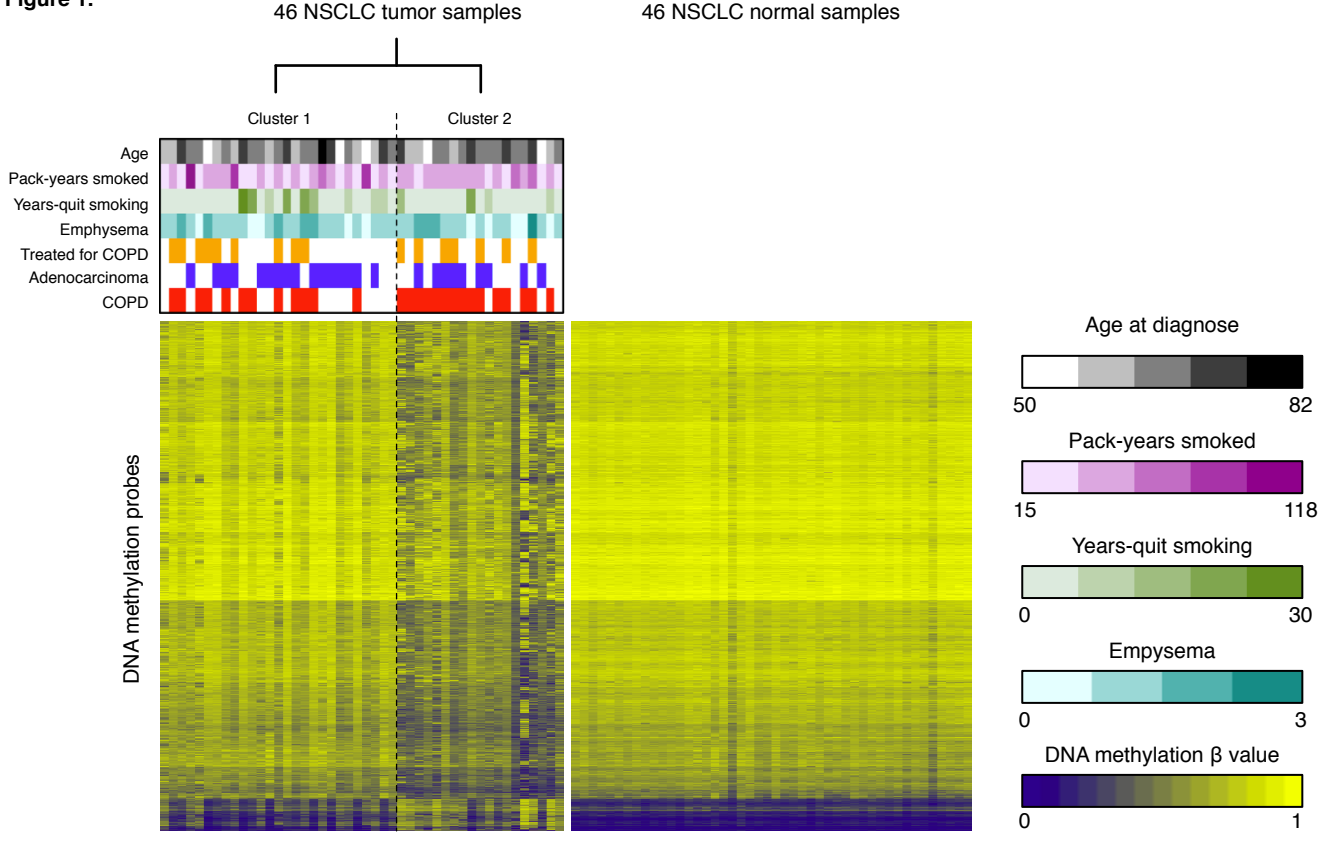
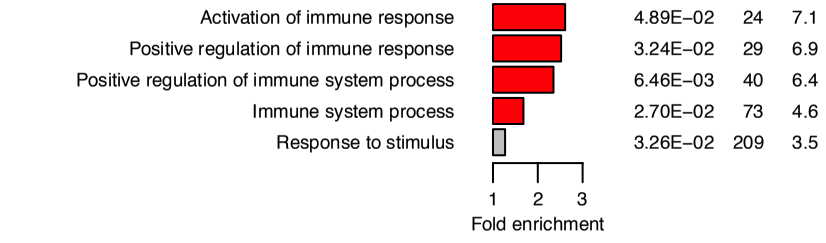

46 NSCLC tumor samples
46 NSCLC normal samples
Cluster 1
Cluster 2
Age
Pack-years smoked
Years-quit smoking
Emphysema
Treated for COPD
Adenocarcinoma
COPD
DNA methylation probes
Age at diagnose
50
82
Pack-years smoked
15
118
Years-quit smoking
0
30
Empysema
0
3
DNA methylation β value
0
1

Figure 2.

Differential methylation between tumor and adjacent lung tissue in the COPD subgroup

A. Tumor-associated hypermethylation



	FDR-adjusted *P*-value	n	%
Activation of immune response	4.89E−02	24	7.1
Positive regulation of immune response	3.24E−02	29	6.9
Positive regulation of immune system process	6.46E−03	40	6.4
Immune system process	2.70E−02	73	4.6
Response to stimulus	3.26E−02	209	3.5

B. Tumor-associated hypomethylattion

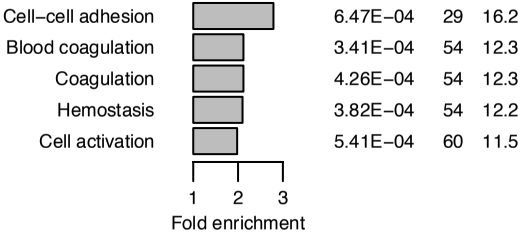
No GO enrichment

Differential methylation between tumor and adjacent lung tissue in the non-COPD subgroup

C. Tumor-associated hypermethylation

No GO enrichment

D. Tumor-associated hypomethylation



	FDR-adjusted *P*-value	n	%
Cell–cell adhesion	6.47E−04	29	16.2
Blood coagulation	3.41E−04	54	12.3
Coagulation	4.26E−04	54	12.3
Hemostasis	3.82E−04	54	12.2
Cell activation	5.41E−04	60	11.5

Figure 3.

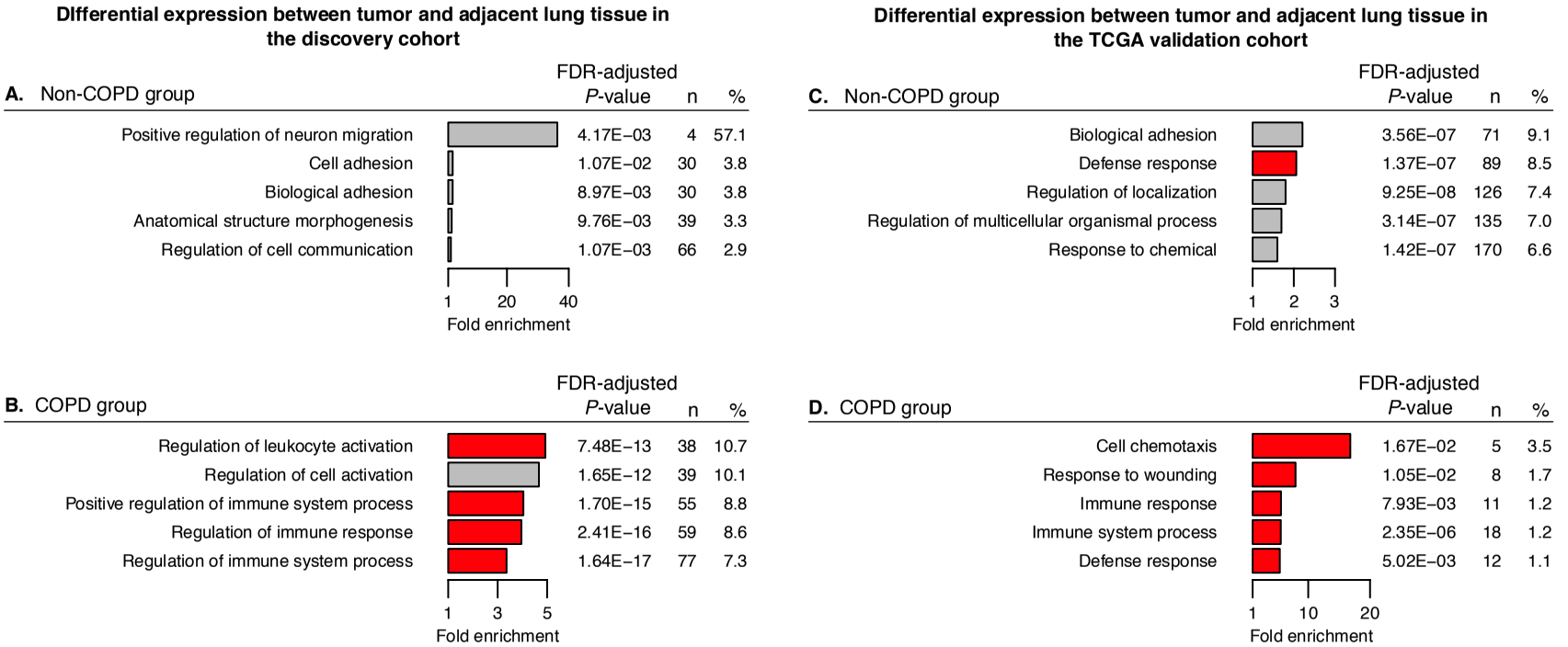
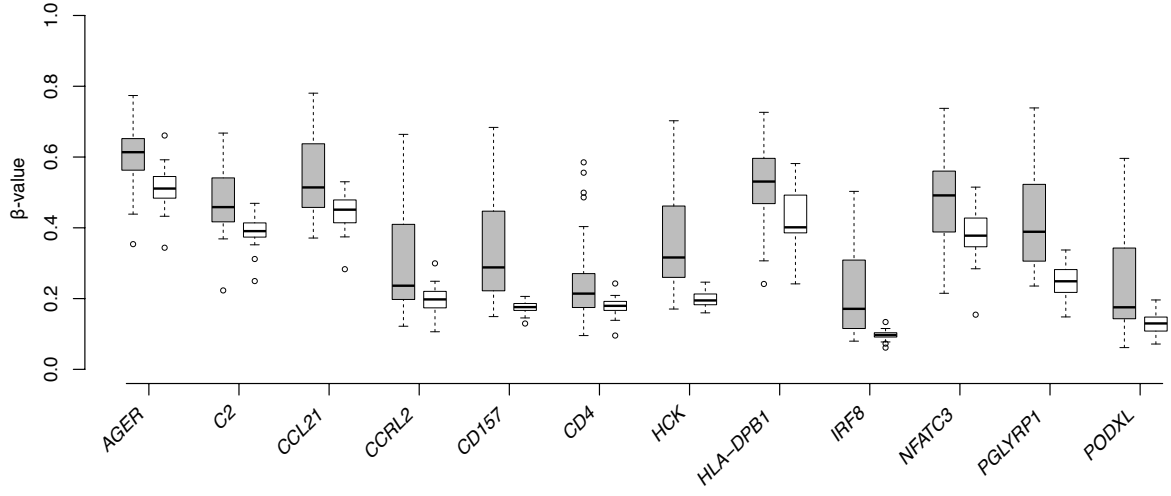

DIfferential expression between tumor and adjacent lung tissue in the discovery cohort
A. Non-COPD group
FDR-adjusted P-value
n
%
Positive regulation of neuron migration 4.17E−03 4 57.1
Cell adhesion 1.07E−02 30 3.8
Biological adhesion 8.97E−03 30 3.8
Anatomical structure morphogenesis 9.76E−03 39 3.3
Regulation of cell communication 1.07E−03 66 2.9
1 20 40
Fold enrichment
B. COPD group
FDR-adjusted P-value
n
%
Regulation of leukocyte activation 7.48E−13 38 10.7
Regulation of cell activation 1.65E−12 39 10.1
Positive regulation of immune system process 1.70E−15 55 8.8
Regulation of immune response 2.41E−16 59 8.6
Regulation of immune system process 1.64E−17 77 7.3
1 3 5
Fold enrichment
Differential expression between tumor and adjacent lung tissue in the TCGA validation cohort
C. Non-COPD group
FDR-adjusted P-value
n
%
Biological adhesion 3.56E−07 71 9.1
Defense response 1.37E−07 89 8.5
Regulation of localization 9.25E−08 126 7.4
Regulation of multicellular organismal process 3.14E−07 135 7.0
Response to chemical 1.42E−07 170 6.6
1 2 3
Fold enrichment
D. COPD group
FDR-adjusted P-value
n
%
Cell chemotaxis 1.67E−02 5 3.5
Response to wounding 1.05E−02 8 1.7
Immune response 7.93E−03 11 1.2
Immune system process 2.35E−06 18 1.2
Defense response 5.02E−03 12 1.1
1 10 20
Fold enrichment

Figure 4.

Tumor versus adjacent lung tissue from COPD patients

A. Increased promoter hypermethylation



B. Reduced gene expression

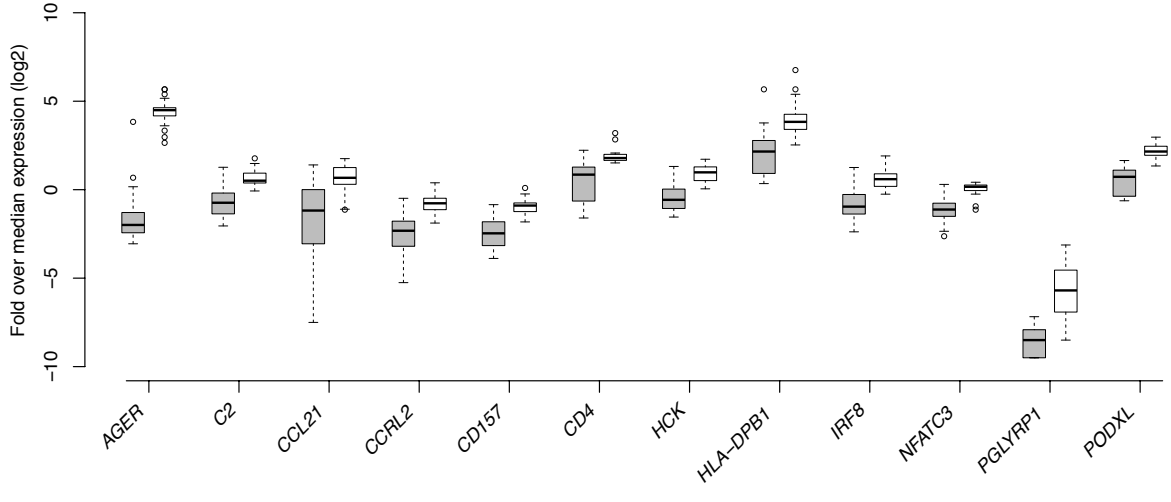
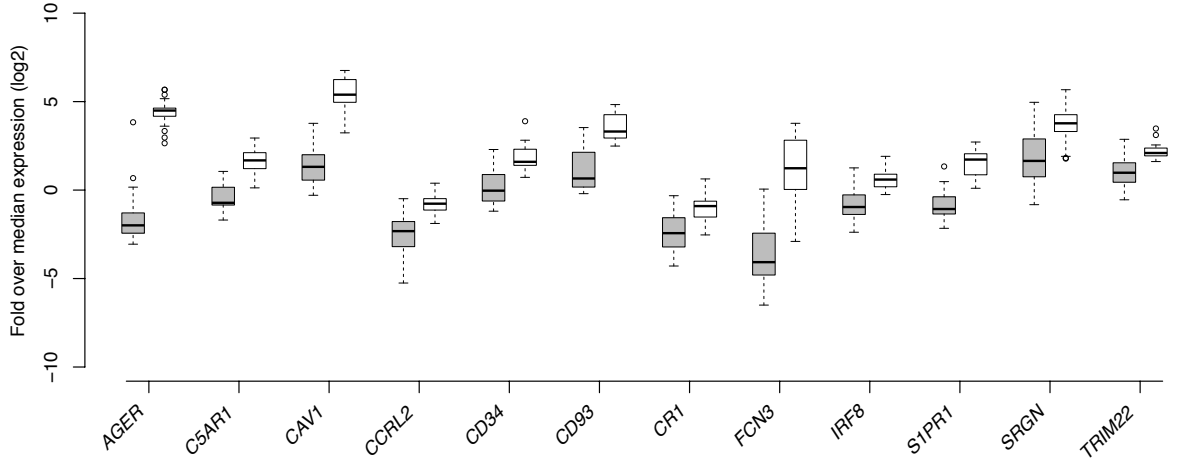


Figure 5.

Tumor tissue from COPD versus non-COPD patients

A. Discovery cohort



B. TCGA validation cohort

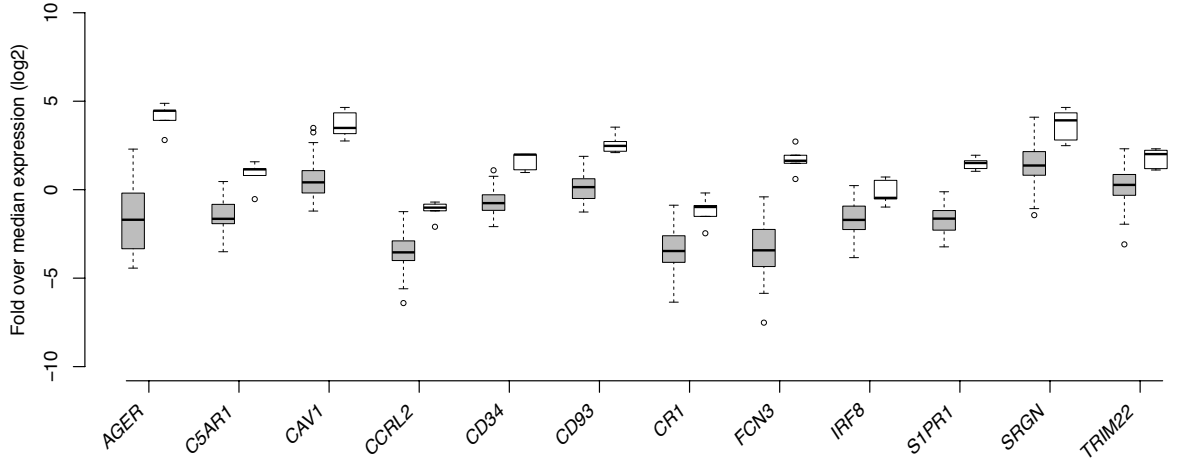
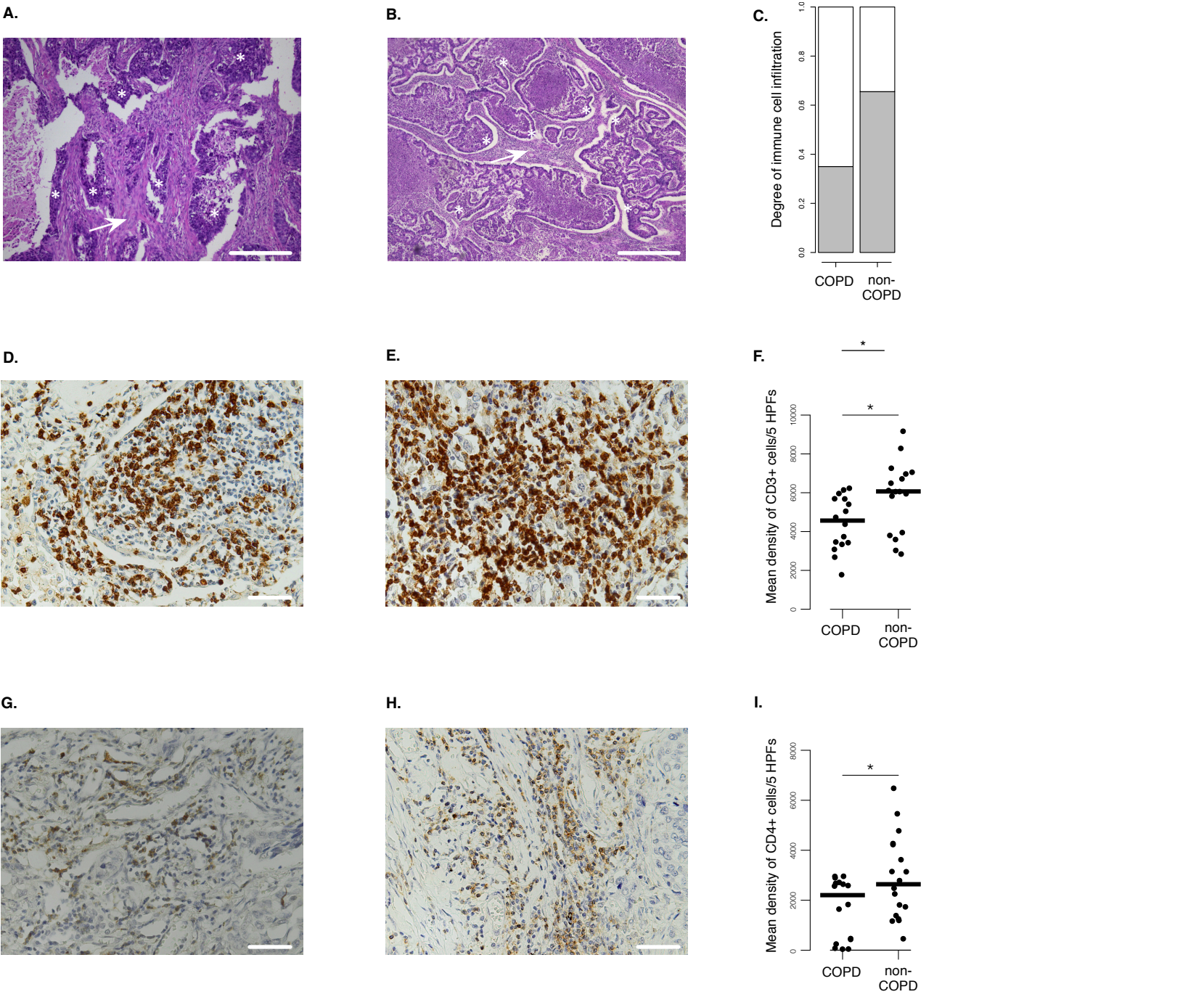


Figure 6.


A.
B.
C.
Degree of immune cell infiltration
0.0
0.2
0.4
0.6
0.8
1.0
COPD
non-COPD
D.
E.
F.
Mean density of CD3+ cells/5 HPFs
0
2000
4000
6000
8000
10000
*
*
COPD
non-COPD
G.
H.
I.
Mean density of CD4+ cells/5 HPFs
0
2000
4000
6000
8000
*
COPD
non-COPD